

DOCUMENT RESUME

ED 406 677

CS 215 802

AUTHOR Pinkerton, K. David
TITLE Enhancing Conceptual Learning by Understanding Levels of Language-Rich Teaching.
PUB DATE Mar 97
NOTE 58p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Processes; Comparative Analysis; High Schools; *Instructional Effectiveness; *Journal Writing; *Language Role; Lesson Plans; Misconceptions; *Physics; *Science Instruction; Small Group Instruction; Teaching Methods

ABSTRACT

Any number of successful teaching strategies use a mixture of methods rich in language use to enhance conceptual learning. Not all language-rich teaching leads to equivalent conceptual gain. A year-long study was conducted in three high school physics classes to provide empirical support for the existence of effects of varying level of language-rich teaching. Three levels (low, medium, and high) of language-rich teaching were investigated. Each level was characterized by a unique language-rich teaching method: low level by hands-on, medium level by small groups, and high level by active mental processing (AMP) journaling (which structures students' dialogue with one another based on compelling demonstrations of everyday misconceptions, prescribes specific notetaking strategies, and requires that students pose an application question each day). First semester posttests revealed that the mean score on assessment instruments for the class using small group methods was higher than the class using hands-on techniques but lower than the class using the AMP journal. All classes were taught in the AMP journal during the second semester. Small group and hands-on classes improved more than the AMP journal class suggesting that high level language-rich teaching intervention is beneficial at any time. (Contains 97 references, and 6 tables and 4 figures of data. An appendix presents sample lesson plans.) (Author/RS)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ENHANCING CONCEPTUAL LEARNING BY UNDERSTANDING LEVELS OF
LANGUAGE-RICH TEACHING

ED 406 677

K. David Pinkerton, Ph. D.

Science Department

Smoky Hill High School

Aurora, Colorado 80015

~~d_pinkerton@shhs1.ccsd.k12.co.us~~

d.pinkert@shhs1.smoky.org

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

K. Pinkerton

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to
improve reproduction quality.
- Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

Submitted 18 Jan 97

Running head: CONCEPTUAL LEARNING AND LANGUAGE-RICH TEACHING

BEST COPY AVAILABLE

25215802

Abstract

Any number of successful teaching strategies use a mixture of methods rich in language use to enhance conceptual learning. Not all language-rich teaching leads to equivalent conceptual gain. A year-long study was conducted in 3 high school physics classes to provide empirical support for the existence of effects of varying level of language-rich teaching.

Three levels (low, medium, and high) of language-rich teaching were investigated. Each level was characterized by a unique language-rich teaching method: low level by hands-on, medium level by small groups, and high level by active mental processing (AMP) journaling.

First semester posttests revealed the order of class means to be: hands-on < small group < AMP journal. All classes were taught in the AMP journal during the second semester. Small group and hands-on classes improved more than the AMP journal class suggesting that high level language-rich teaching intervention is beneficial at any time.

Enhanced Conceptual Learning by Understanding Levels of Language-rich Teaching

Good lessons are an excuse for students to talk. But not all talk is the same. Education has experienced a major shift in the methods used to teach important concepts. The shift from telling students concepts to students generating concepts themselves, in highly interactive social environments, has become pervasive (Dufresne et al., 1996). Language has always been the primary agent of social interaction that utilizes concepts (Bialystok & Hakuta, 1994; Vygotsky, 1978) and language-rich teaching is at the center of improved teaching techniques developed over the last 10 years.

A spate of information about language-rich teaching methods has been published in the last decade. Techniques have been developed for the entire range of content areas, ages, and ability levels. Whether reciprocal teaching (Brown & Palincsar, 1989), scaffolding (Brown & Clement, 1991), small groups, (Cohen, 1994; Webb, 1989), tutoring (Cohen, Kulik, & Kulik, 1982; Graesser, 1992), or reciprocal questioning (King, 1990), the core of the method involves students talking and writing about important ideas.

In the domain of physics, the effort to create language-rich classrooms has been especially prodigious and the results most pronounced. Socratic Dialogue (Hake, 1992), microcomputer-based laboratory (Thorton & Sokolov, 1990), coaching (Ludicio, 1992), questioning (Minstrell, 1991), cognitive conflict (Mazur, 1993; Meltzer & Manivannan, 1996), modeling (Halloun, 1987), workshop physics (Laws, 1991a), and hypermedia (Wilson, 1994) all represent language-rich methods that increase students' conceptual gains by fostering language interaction that draws from the concepts at hand. Hake(1994) contrasted language-rich physics teaching methods (which he labeled *interactive engagement*) with traditional physics teaching methods in over 50 quantitative studies that used the same test of conceptual gain - the Force Concept Inventory (Hestenes, Wells, & Swackhammer, 1992). He predicted gain scores as a function of pretest and

found that language-rich methods yielded $r = .51$ and traditional methods yielded $r = .24$. In his analysis, language-rich teaching seemed to enhance conceptual learning.

Common Features of Language-rich Teaching

There are a number of features common to quality language-rich methods mentioned. (In context, rich does not mean more, rather it refers to the quality of the language experience.) Quality manifests itself in language-rich teaching strategies in four ways. First, students actively generate meaning through language. Although the mechanism of how language captures concepts varies from a neuro-evolutionary view (Edelman, 1992), to a cognitive science perspective (Cain & Cain, 1991; Pinker, 1995), to a social cognitive view (Bandura, 1986; Vygotsky, 1978), there is little dispute that language and conceptual structures build together. Wittrock (1990), building on the earlier experimental work of Slamecka and Graf (1978), proposed a generative model of comprehension to explain why small group interactions foster higher level cognitive development. Language is a generative activity because it generates multiple relations and representations, requires recall from knowledge and experience, generates summaries and analogies, constructs relevant representations, fosters metacognition, and adapts to analytical and holistic thinking.

Second, students elaborate on new concepts through language. Verbal explanations of conceptual details can overshadow visual input (Clark & Pavio, 1991; Mayer & Sims, 1994). This suggests that language-rich elaborations have more impact on conceptual learning than visually encoded information. When teaching methods encourage students to elaborate their conceptual understanding, students generate and rehearse conceptual details and features, create parallel connections or relationships to other conceptual structures, and develop labels and cues that foster memory and automatization (Pressley, et al, 1987; Simpson, 1994). Elaboration has played an

important part in improved achievement for students in graduate education (King, 1990), elementary mathematics (Webb, 1991), and college chemistry (Prestel, 1993).

Third, language-rich teaching methods include specific strategies to help students acquire, modify, and habituate concepts. Students require active strategies to develop reasoning involving new concepts learned in school (Sere, 1991). Both the strategies and their instruction require highly interactive dialogue. Examples of specific cognitive strategies utilizing language-rich techniques vary in application from elementary school reading comprehension (Davey & McBride, 1986) to college physics laboratory comprehension (Hake, 1992).

Fourth, language differentiates between the conceptual structures of experts and novices. The way students talk about concepts reveals their position on a novice to expert continuum. Novices often scan instructional events for novel features and mesh them into their existing conceptual framework by memorizing conclusions (Minstrell, 1991). Experts seek and come to understand the underlying conceptual structure of situations (Chi, Glaser, & Farr, 1988). Once teachers assess students' degree of expertness, proper interventions can be implemented to enhance achievement. A number of review articles detail the teaching strategies and learning environments necessary for enhanced conceptual learning predicated on teachers being able to distinguish between novices and experts (Carey, 1985; Gilbert & Watts, 1983; Scott, Asoko, & Diver, 1991; Pines & West, 1985). Several authors implicate language use as a way to nurture expert-like behavior in students (Chan & Burtis, 1992; Chi et al, 1989; Chi & VanLehn, 1991; Cohen, 1994).

Differences in Language-rich Teaching

While there are many similarities, a difference exists among language-rich teaching methods. This difference depends on the richness of teacher-student language interaction. The notion of richness is founded in Vygotsky's zone of proximal

development (Howe, 1996; Vygotsky, 1978; Whelan, 1994). Language functions as a mediator between everyday (naive) and scientific (systematic) conceptualizations of the world. Everyday concepts emerge spontaneously from innate components in humans and are reinforced by familial interaction. Scientific concepts result from linguistic interaction with those more knowledgeable. Scientific concepts are the primary concern of school. They represent concepts that would, in large part, not be learned unless students went to school. Richness is measured by the extent to which a teaching method fosters discourse between a student (novice) and a teacher (expert). Teaching methods that foster more student-teacher dialogue have an increased chance of using language to help students learn school concepts.

Like learning a second language, conceptual learning in a new domain is facilitated by linguistic interaction with experts in that field. Language and conceptual structures build together because language and meaning influence each other (Bialystok & Hakuta, 1994). The more high quality discourse that students can have with those more knowledgeable in a domain the greater likelihood that students will learn school concepts.

Increased achievement is enhanced when student-teacher interaction is structured. For example, Cohen, Kulil, and Kulik (1982) reviewed 65 tutoring programs in which students served as tutors and found a mean effect size of .40 with the largest gains occurring in mathematics and low achieving students. Rosenshine and Meister (1994) reviewed 18 reciprocal teaching programs and found a mean effect size of .32 on standardized tests of reading comprehension. Perhaps these results are due to the increased natural dialogue between students and teachers or trained tutors. As Graesser (1992) noted, students in structured tutoring conditions ask 100 times more questions than students in traditional settings. Clearly, tutoring and reciprocal teaching provide a different level of language-rich interaction between teachers and students.

Richness can vary. This is especially evident in science classrooms. Common language-rich science teaching methods such as hands-on experiments and small group problem solving work differ in richness of language interaction. For example, talking about laboratory experiments is certainly different than listening to lectures. Yet when lab team members talk to each other, their conversations focus on the objects of the experiment--how to finish on time or how to get the expected answer--not the underlying school concepts (Hodson, 1993). Though students are engaged in much dialogue in hands-on settings, they tend to focus on the surface features of immediate everyday objects depicted by lab apparatus rather than the deeper conceptual implications which actually govern the outcome of the experiment.

A possible explanation for students' attention to the commonplace and expedient features of hands-on activity is noise (Newman, 1985). Noise results from students being asked to process too many things at once. Having to understand the nature of the experimental problem as well as read, comprehend, and follow experimental instructions; interpret results; write an account of the activity; and get along with partners creates noise that erects barriers to learning. Because of the cognitive load (Sweller & Chandler, 1991) associated with traditional hands-on activities, students' dialogue revolves around concrete aspects of the laboratory rather than conceptual issues. In fact, some research suggests that hands-on activities are no better than lecture-discussion at improving scientific knowledge (Gunstone & Champagne, 1990; Novak, 1990; Tobin, 1990), and are counterproductive to students' understanding of scientific inquiry (Klopfer, 1990). Hands-on activities get students talking but represent a low level of language-rich teaching.

Small group work is a prominent feature of many science classrooms (Hake, 1992; Laws, 1991b; Lumpe & Staver, 1995). In a pragmatic effort to take advantage of the conceptual gains associated with language-rich interaction among students, without having to dialogue with every student for extensive periods of time, teachers use small

groups. When small groups are employed in science classes, gains in conceptual understanding and applications occur (Cohen, Lotan, & Leecher, 1989; Hake 1994), ability to generate rules increases (Schwartz, Black, & Strange, 1991), and improvements in attitudes, attendance, and participation occur (Cottel & Lundsford, 1995). Shachar and Sharan (1994) investigated Israeli junior high school students' verbal behavior, social interaction, and academic achievement. In all areas, small group classes produced more desirable behaviors than traditional classrooms. The authors attribute these findings to speech strategies that cause students to focus on words as carriers of meaning. Students restructure their communication in light of their own thoughts in order to gain control over the verbal symbols and the kind of message verbal symbols express to their peers.

The language-rich quality of small groups depends on group composition. When low and high achieving students are grouped, low achieving students improve on measures of factual recall, application, and problem solving (Trudge, 1990). Yet heterogeneous high achieving groups do not improve on similar measures (Hooper & Hannafin, 1988). This suggests that working with more capable peers in the language-rich environments of small groups enhances achievement especially when the interaction is with more capable peers. These findings exemplify small cooperative groups' utilization of the *zone of proximal development*. More knowledgeable peers foster movement toward acquisition of school concepts in less capable students by using rich language interaction. But because these peer interactions are not at the same level of richness as interactions with teacher-experts, small groups represent a middle level of language-rich teaching.

Though individual teacher-student dialogue would represent the highest level of language-rich teaching, it is not practical to expect a teacher to individually tutor 25 students. Economic reality dictates that in the absence of one student per teacher, other high-level language-rich teaching strategies should be developed. Student journals, if properly applied, can produce many of the same effects as one-on-one interaction.

Journaling that is effective in promoting students' acquisition of systematic school concepts requires strong leadership. Teachers must have active strategies to draw students' attention to reasoning and cause them to express everyday concepts in contrast to formal school concepts (Sere, 1991). Grumbacher (1987) exhibits her strategies for using journals in high school physics: (a) The best problem solvers are able to relate physics to their personal experiences; (b) Writing helps students find connections between experience and theory; (c) Students will do more work when they are seeking answers to questions that they initiate; (d) Journals foster such questions; and (e) Students need time to think about new ideas. But a delicate balance between structure and open-endedness must be achieved. If teachers do nothing to structure the level of interaction, they may well find that students stick to the most concrete mode of interaction. If teachers do too much to structure interaction, they may prevent students from thinking for themselves and gaining benefits from the interaction (Cohen, 1994, p. 22).

The active mental processing (AMP) journal (Pinkerton, 1996) may be a tool that science teachers can use to achieve the balance that Cohen mentions. This journaling method depends on teacher intellectual leadership and students who are motivated to understand school concepts. It structures students' dialogue with one another based on compelling demonstrations of everyday misconceptions; prescribes specific note taking strategies (Chandler & Sweller, 1992; King, 1992); and requires that students pose an application and a question each day. Teachers can "shrink" classroom size by training students to pose teacher-like questions of their peers, then teachers monitor conversations around the room to diagnose inappropriate use of key concept words. In this way, the AMP journal fosters high level language-rich teaching without continual one-on-one, student-teacher linguistic interaction.

Research has converged to say that language-rich teaching methods promote conceptual learning more than traditional didactic methods. The study herein represents a natural extension of research into language-rich teaching methods. Rather than

suggesting that all language-rich methods produce equivalent conceptual gains in students, it professes that levels of language-rich teaching exist. In this study of science education, the levels are represented by hands-on (low level), small groups (middle level), and AMP journal (high level) methods. If differences in levels of language-rich teaching exist, then differences in students' conceptual learning should be detected, with the highest level of language-rich teaching producing the greatest conceptual gains. An AMP journal class (high level) should demonstrate greater conceptual understanding than a small group (medium level). A hands-on (low level) class should demonstrate less conceptual comprehension than the small group class. Further, if small group and hands-on classes are switched to the AMP journal method after several weeks, then they will experience greater conceptual gains than the AMP class.

Method

Design

This study utilized a quasi-experimental, non-equivalent cohort design partitioned into three levels of treatment to foster internal validity (Cook & Campbell, 1979). Three college preparatory high school physics classes, not individual students, were randomly assigned a treatment level. This type of design carries with it several validity concerns. First, teacher by treatment confounds threaten internal validity (Lysynchuk et al., 1989). Therefore, the teacher-researcher instructed each class in this study.

Second, to detect teacher bias for a particular method, two measures were taken. At the end of the first 18 weeks of the study (one semester), a 25 question Likert-scale survey was given to all 79 students to assess differential teacher enthusiasm. No statistically significant differences were detected on the total score, $F(2, 76) = 2.33$, $p = .10$, with the order of increasing means (more enthusiastic) being AMP journal < hands-on < small group classes. Also, two administrators who were trained in teacher evaluation observed classes on a random schedule. They completed identical 12 question

Likert-scale surveys designed to monitor teacher enthusiasm. No statistically significant differences were detected among classes, $F(2, 9) = .50, p = .63$.

Third, differential exposure to instruction can invalidate findings. Careful experimental controls were used to ensure that differences in conceptual gain could be attributed to the level of language-rich teaching, not different instructional materials, hand-outs, or tests. All classes used the same textbook, were assigned the same homework and readings, performed the same formal laboratory experiments, solved the same chapter problems, received the same handouts, and were asked the same questions during teacher presentations. Fourth, strict measures were enacted to ensure that the amount of time spent by students on treatment and non-treatment tasks was equivalent. A nine question Likert-scale survey was administered at semester's end to detect any differences in homework time. No statistically significant differences were detected among classes, $F(2,76) = .09, p = .92, M = 3.2$ hours/week, $SD = 1.3$ hours.

Fifth, process measures document what students are actually doing and can strengthen conclusions about instructional effects (Lysynchunk et al., 1989). Two process measures were used. (1) Unit tests covering from one to three chapters were designed to monitor implementation of procedures. Each test featured three questions with multiple parts that drew from attributes of each treatment level not associated with conceptual development. For example, one question on each unit test asked students to perform tasks associated with hands-on activities such as measuring angles or devising a procedure. Questions for the AMP journal and small group classes were included in each of four unit tests. Significant differences among classes on their respective process measure question would suggest that process were performed as intended.

Several students' tests were randomly selected, scored by the teacher-researcher and another physics teacher (interrator reliability = .81). A two way (period x gender) ANOVA was also performed. The AMP journal students performed statistically better on their process measure question than the small group or hands-on class, $F(2, 79) = 5.80$,

$p < .01$, power = .87, $\eta^2 = .14$. No other statistically significant differences were detected. (2) A daily teacher-activity log was kept. Entries focused on any unanticipated differences among classes and what students actually did relative to what was planned.

Sixth, mono-operational bias can lead to lower validity since single operations underrepresent constructs and contain irrelevancies. To combat this problem, two methods of monitoring students' conceptual growth were used, a traditional 29 question multiple choice test named the force concept inventory (FCI) and a conceptual configuration (concept map) measure (NETSIM). These measures use vastly different approaches to monitor conceptual learning.

Seventh, initial group differences can make it difficult to demonstrate treatment effects (Tabachnik & Fidell, 1989). In lieu of random assignment of subject to treatment level, three devices were used to account for potential group differences before instructional treatment: (1) Covariates were used to match groups statistically before treatment. For example, pretests on FCI and NETSIM were used as covariates in subsequent ANCOVAs. (2) Strict control over extraneous variables helps to ensure that differences on dependent measures are due to treatment not factors such as those mentioned as validity concerns. (3) Demographic data such as age, gender, number of years of math, number of years of science, and overall grade point average were collected. These data are summarized in the participants section.

Participants

This study was conducted in a large (2,500 students) middle to upper middle class suburban high school in a large mountain states city. The school ethnic mix was 1% Hispanic, 3% Asian, 9% African American, and 87% Anglo. Sixty-eight percent of students take more than the required two years of science. There are typically ten sections of physics per year with approximately 25 students per section. A summary of

demographic data is found in Table 1. A series of one way ANOVAs detected no statistically significant differences among classes on any demographic variable.

Insert Table 1 about here

The teacher-researcher had 18 years of teaching experience, 11 years was in physics. The teachers who were used to create the teacher-expert composite concept map came from the same school district and had similar backgrounds as the teacher/researcher.

Instruments

Force Concept Inventory (FCI) -- The FCI is a 29 question multiple choice test of beginning concepts in physics (Hestenes, Wells, & Swackhamer, 1992). It is widely used to diagnose misconceptions in physics and to evaluate instructional effectiveness. The authors suggest that preconditions - such as students' mathematics background, socioeconomic level, and teacher's experience - have little effect on gain scores. This suggests that the FCI can be used in a wide variety of circumstances to monitor teaching methods that foster conceptual understanding.

The Kuder-Richardson reliability has been found to be .86 for the pretest and .89 for the posttest (Halloun & Hestenes, 1985). For this study it was .54 for the pretest and .76 for the posttest. Face and content validity were determined by (a) extensive review by physics professors and graduate students, (b) preliminary testing with graduate physics students until all agreed on the answers, (c) interviews of introductory physics students who had taken the test to check for understanding, and (d) detailed analysis of student scores to find evidence of common misunderstandings. A gender by level two way ANOVA revealed no initial statistically significant differences among treatment levels on

the FCI pretest, $F(2, 79) = .26$, $p = .77$, but an expected significant gender effect, $F(1,79) = 19.6$, $p < .001$.

Factor analysis has been performed to detect multidimensionality. Huffman and Heller (1995) detected two weak factors that combined to explain 15% of the variance. A Rasch item analysis was performed in this study and no strong evidence for multidimensionality was detected. On balance, the FCI seems to monitor introductory physics concepts effectively.

NETSIM -- Relatedness measures, which are ultimately converted to NETWORK SIMilarity (NETSIM) scores, monitor conceptual learning in a vastly different fashion than the FCI. A computer algorithm called Pathfinder networks (Schvaneveldt, 1990) randomly selects two words from a list supplied by the experimenter, e.g. acceleration and gravity. Students move a cursor between *high* and *low* to indicate their sense of relatedness between the two concept words. All students rated 106 possible combination of words.

Pathfinder generates concept maps by making links based on relatedness scores. Highly related concepts are directly linked and less related concepts are separated by two or more links. The resulting concept map captures the configural character of domain knowledge and thus represents the most salient relations among concepts (Gomez, Hadfield, & Housner, in press).

A composite concept map using relatedness scores from 10 physics teachers was generated by the Pathfinder program. All teacher's relatedness scores correlated significantly with every other teacher. The lowest r was .35 between the oldest and youngest teacher. The highest r was .69 between this researcher and his officemate.

Gomez, Hadfield, and Housner (in press) describe how the NETSIM score is calculated. The observed similarity between a student's concept map and the teacher-expert composite is calculated by dividing the number of links shared by both networks by the number of links in either network. The expected similarity is calculated using the

probability that two networks share x links. The NETSIM score results from observed similarity minus expected similarity. NETSIM values greater than zero indicate a greater degree of similarity between two networks than is expected by chance. NETSIM values less than one suggest that the observed network similarity is less than expected by chance. Students who experience greater conceptual learning can be expected to have larger positive NETSIM values.

The Pathfinder approach has been found to be valid in discriminating between expert and novice conceptual structures (Durso & Coggins, 1990), delineating pilot's conceptual structures for use in designing flight simulator controls (Roske-Hofstrand & Paap, 1986), and predicting examination scores (Goldsmith, Johnson, & Acton, 1991).

Learning Styles Inventory (LSI) -- The LSI (Kolb, 1985) evaluates how a person prefers to learn and how he or she deals with day-to-day situations. It features 12 sentence completion items which allow scoring along four dimensions of learning-- concrete experience (CE), reflective observation (RO), abstract conceptualization (AC), and active experimentation (AE). Two learning style scores are derived from the four learning dimensions. The AC-CE score is calculated by subtracting the concrete experiencing value from the abstract conceptualization value. The AE-RO score is calculated from subtracting the reflective observation value from the active experimentation value. Cronbach's alpha for the AC-CE and AE-RO subscales were .86 and .87 respectively. Intercorrelations were used to determine that the subscales were orthogonal as intended.

Procedure

Procedures were devised and implemented to ensure that level of treatment was the only systematic difference among classes. Within the first three days of class, grading procedures, policies, student information sheet for demographic data, FCI pretest, relatedness premeasures, and LSI scores were administered. Treatment ensued and fell

into a two- to four-day pattern for coverage of new concepts. For the small group and hands-on classes, the pattern was teacher demonstration, teacher led discussion, teacher explanation, and quiz. Formal laboratory experiments, reading and problem assignments, and unit tests were identical for each class and were not part of this pattern. Quizzes for the hands-on class lasted one class period and involved physical manipulation of laboratory equipment focused on an experimental task such as determining the acceleration of a ball down a ramp. Teams of three or four students wrote theory, hypothesis, procedure, observation, and conclusion sections for the quiz. Quizzes in the small group class consisted of teams solving assigned chapter problems using a prescribed problem solving template.

AMP journal students experienced the same teacher demonstrations and teacher -led discussions, but were led by the teacher to make programmed entries into their journals. These entries were punctuated with pre and postdictions, in which the teacher prompted correct use of concept words use as students recorded their thoughts. Students were asked to verbalize their journal entries to peers and make adjustments in their journals if classmates did not understand. Each day students were required to create and answer a *question of the day* and to record an *application of the day*. Emphasis was placed on applications that students drew from their personal experience and history. If students could not think of an appropriate application of physics for the day, they were allowed to create a metaphor that helped them understand a difficult concept. A sample AMP journal lesson can be found in the appendix.

AMP journals were collected every three weeks and graded for completion and procedure rather than accuracy. Optimum points for quizzes equaled those of the journal. All other graded assignments were identical for each class.

At semester's end (18 weeks), postmeasures of relatedness scores were collected and the FCI was given as a final. After all tests were completed, the student evaluation of teacher enthusiasm was administered. On the first day of the second semester, students

were shown raw FCI mean scores per class. The teacher informed students that all classes would be taught by the AMP journal method from that point on, but that formal submission of the journal would be voluntary. Full compliance would yield approximately 125 journals per class per semester. The numbers of journals submitted during the voluntary system per class were: AMP journal = 47, small group = 30, and hands-on = 22. Topics for the first semester were addressed only if they became important in explaining second semester concepts. In week 36, the FCI was given as part of the second semester final.

Results

Results can be divided into two parts: (1) first semester analyses of FCI and NETSIM data using ANCOVA, (2) and second semester FCI data using repeated measures ANOVA.

First semester analyses

FCI data were screened for entry accuracy, analyzed for outliers, and checked for assumptions required by ANCOVA. Two students dropped physics in the first week of class and one student did no homework and failed to take the final. These two students were not included in analyses. No univariate or multivariate outliers were detected in the pretests, while two univariate outliers were detected in the posttests. Both cases were in the hands-on class. One was the highest score (100% correct) and the other was the lowest score (21% correct). Both cases were retained.

ANOVA assumptions include independence, normality, and homogeneity of variance. Scores were independent, since teacher monitoring ensured that all students were tested on an individual basis. No skewness or kurtosis values were greater than ± 1.0 and scatter plots of standardized residuals revealed no unusual patterns. Therefore,

normality was assumed. A Levine statistic for the posttest of 1.24, $p = .29$ suggested no significant violation of homogeneity of variance.

The FCI pretest was used as a covariate and required that two additional assumptions be met. MANOVA revealed no significant treatment level by pretest interaction, $F(2, 79) = .26$, $p = .77$. Thus, homogeneity of regression was reasonably met. Cronbach's alpha was .54 for the pretest. This indicates error in measurement of the covariate. This issue is addressed in the limitations section.

Descriptive statistics, as well as omnibus and planned contrasts F tests, are summarized in Tables 2 and 3, respectively. Adjusted means per gender and class period, as well unadjusted means for each treatment level, are show in Figure 1.

Insert Table 2 about here

Insert Table 3 about here

Insert Figure 1 about here

The omnibus gender by treatment level interaction and the main effect of treatment level were significant at $p = .06$ and $p = .05$ respectively. Planned contrasts were used to probe interaction and treatment level main effects. The first contrast compared the adjusted means of the AMP journal class with a combination of small group and hands-on classes, $p = .01$, effect size = 1.25. The AMP journal adjusted class mean on the posttest was highest. The adjusted means between the small group and hands-on classes were not significantly different. The second contrast probed the gender by treatment level interaction between small group and hands-on classes only, $p = .03$,

effect size = .60. Females in the small group class unexpectedly scored lower than females in the hands-on class even though the adjusted mean for the small group class as a whole was higher than the adjusted mean for the hands-on class.

The same analyses were performed using the AC score from the LSI measure. Pretreatment measurement of AC indicated a gender difference but not a class period difference. The contrast between the adjusted means of the AMP journal class and a composite of small group and hands-on classes was significant, $F(1, 79) = 5.73, p = .02$. These results mirror those obtained with the FCI pretest used as a covariate.

NETSIM data were analyzed in the same manner as FCI data. NETSIM premeasures were used as a covariate to the dependent variable--NETSIM post measure. The 2 x 3 factorial ANCOVA required appropriate data screening and assumption testing. Pre- and posttreatment kurtosis were 3.48 and 1.75, respectively, and pre- and post-treatment skewness were 1.44 and .95, respectively, but no transformations of the data were performed because ANCOVA is robust with regard to violations of normality. All four univariate outliers and both multivariate outliers were high scores. They were retained in analyses. No violation of homogeneity of variance was detected by the Levine statistic, $p = .40$. MANOVA revealed no statistically significant interaction between pre NETSIM and treatment level therefore, homogeneity of regression was assumed for the covariate.

The pattern of results for NETSIM data was extremely similar to results for FCI data. Factorial ANCOVA (gender x treatment level) results for NETSIM data are summarized in Tables 4 and 5.

Insert Table 4 about here

Insert Table 5 about here

Once again, omnibus F tests found a statistically significant difference among levels of language-rich teaching treatment with the order of means being as hypothesized: AMP journal > small group > hands-on. A significant difference was found between the adjusted means of the AMP journal and a composite of small group and hands-on classes, $p = .01$, effect size = 1.26. Adjusted means between the small group and hands-on classes were not significantly different. No statistically significant gender-by-level interaction was found in the small group and hands-on classes as occurred with the FCI results, but the pattern of interaction was identical as depicted in Figure 2.

Insert Figure 2 about here

A variety of correlations were calculated to investigate relationships between students' performance on the FCI and NETSIM (concept maps). Pre- and postmeasures of NETSIM correlated at $r = .35$ while pre- and posttests for the FCI correlated at $r = .86$. Postmeasures of NETSIM and FCI posttests correlated at $r = .39$. This suggests that FCI and NETSIM capture different aspects of conceptual learning. Support for this notion follows from analyses with the raw relatedness scores. These scores were not manipulated by the pathfinder algorithm which creates a concept map of students' beliefs. Students' pre- and post raw relatedness scores correlated at $r = .91$. ANCOVA analyses with these data revealed no reliable differences among classes. Thus, conceptual learning differences were not detected. Creating a configural measure of students' conceptual

position with pathfinder seems to be more effective at detecting conceptual learning than raw relatedness scores.

Second semester

The FCI was given as part of the second semester final and repeated measures analyses performed on the pretest and two posttests. This was done to investigate whether application of high level of language-rich teaching--to groups previously taught by medium or low level methods -- experience greater conceptual learning when switched to a higher levels of language-rich teaching. Further, effects of high levels of language-rich teaching on retention and transfer could be investigated.

No univariate or multivariate outliers were detected and no violations of normality or linearity detected for the second posttest. A Levine statistic of 3.39, $p = .04$ indicates a violation of the homogeneity of variance assumption for the second posttest. According to Glass and Hopkins (1984), actual $\alpha = .06$ with a nominal $\alpha = .05$ for the reported variances calculated for posttest 2. The homogeneity of covariance assumption for repeated measures was met since Box's $M = 12.1$, $p = .50$ and no large violation of sphericity was detected as indicated by a Greenhouse-Geiser epsilon, $\epsilon = .86$.

MANOVA was used in a mixed (within and between subjects) analysis of the three FCI tests. The between subjects factor was level of language-rich teaching and the repeated factor was FCI test. Only univariate tests are reported.

Contrasts were used to investigate differences in conceptual learning among treatment levels from the first to the second posttest as measured by the FCI. A statistically significant difference among treatment levels was still evident with the order of means unchanged from the first semester data, $F(2,79) = 3.81$, $p = .03$. A difference in the two posttests was detected significant at $p = .11$, $F(1, 79) = 2.61$. A simple effects contrasts exposed a pattern of differential increase on the FCI score. That is, though all classes scored higher on the second posttest, the small group and hands-on classes

improved more than the AMP journal class. To probe this effect further, interaction of contrasts was utilized.

Two between subjects contrasts were tested against two within subjects contrasts (Levine, 1991). The between subjects contrasts were identical to those used with the first semester FCI data. These contrasts compared group differences between the AMP journal class and a composite of small group and hands-on classes, as well as between the small group and hands-on classes only. The within subjects contrasts followed the same pattern but investigated differences among tests. The first repeated measures contrast compared the pretest to a composite of two posttests, and the second contrast compared the two posttests.

A significant interaction between the within and between subjects contrasts was detected. Between subjects contrast one, AMP vs. (SG + HO), interacted significantly with the first within subjects contrast, pretest vs. (posttest 1 + posttest 2), $F(2, 79) = 4.5$, $p = .04$. This result suggests that the AMP journal class maintained its superior performance on the FCI test averaged over two posttests. A significant gender by level by test interaction for the two posttests was obtained, $F(2, 79) = 6.44$, $p < .01$. That is, males in the hands-on and females in the small group classes improved from posttest 1 to posttest 2 while females in the hands-on and males in the small group produced lower scores on the second posttest. Figure 3 depicts class means for the two posttests graphically.

Insert Figure 3 about here

Discussion

First semester results of differential conceptual learning measured by FCI and NETSIM tell the same story, but from different perspectives. The message of the story is that when different levels of language-rich teaching are applied, differential conceptual

learning occurs. In other words, effects of adjusting levels of language-rich teaching exist. Furthermore, a hierarchy of language-rich teaching levels exist in which high levels produce the greatest conceptual learning in students.

Other studies support this conclusion. Using path analysis, Pizzini and Shepardson (1992) compared two groups of eighth-grade biology students on behaviors such as attending, responding, following, and soliciting. Groups solved problems using traditional laboratory or small group methods. Students' behavior in the small group class correlated significantly with lesson structure, while those in the traditional hands-on class did not. This results suggests that at least two levels of language-rich teaching exist and that small group interaction produces student performance more closely aligned with teachers.

Howe et al. (1995) studied four types of peer collaborative groups in middle school physical science. Their results suggest that small groups, which were exposed to hands-on and high level language-rich teaching, produce greater conceptual learning than hands-on only. This suggests that high levels of language-rich teaching add cognitive resources to hands-on methods that result in enhanced performance. The current study investigated differential conceptual learning partitioned along three levels of language-rich teaching in order to understand better its full effect.

An increasing number of theorists are evoking a Vygotskian perspective in science education (Bowen & Roth, 1995; Lemke, 1990; Howe, 1996; O'Loughlin, 1992). Vygotsky's perspective suggests that language mediates students' intellectual progress from every day to school concepts. The cognitive science view utilizes, on the other hand, four aspects of language: lexical, morphological, sentence, and discourse levels (Caplan, 1992). Both perspectives converge to explain why a high level language-rich treatment produced reliable increases in conceptual learning.

Vygotskian and cognitive science explanations depend on differentiating between concept types. Concepts typically broached in school tend to be abstract. Learning them

requires extensive, programmed verbal interaction with teacher-experts. Everyday concepts are primarily object oriented and to a greater extent can be encoded visually. This multimodal view of semantic processing delineates between visual and verbal concepts (Caplan, 1992; Clark & Paivio, 1991). If verbal concepts are more abstract than object concepts then language-rich interaction should foster learning school concepts and common experience will reinforce everyday concepts. In effect, when learning abstract school concepts is deemed important, both Vygotskian and cognitive science views suggest a hierarchy that supports language-rich teaching to enhance school concept learning.

This hierarchy manifests itself in the FCI and NETSIM results of the current study. The FCI reflects more of the morphological and sentence-level aspects of language's role in conceptual learning. This is due to the structure of the multiple choice test. It uses full sentences constructed around variations of key concept words in a static document. NETSIM measures the lexical dimension of concepts using techniques very similar to multidimensional scaling. NETSIM probes conceptual understanding at the word level by mapping conceptual configurations built from relationships among concept words. Jointly, FCI and NETSIM present a more complete picture of conceptual learning because together they tap into a broader representation of language. Results from both measures indicate that high levels of language-rich teaching surpass medium and low levels when learning school concepts is the goal.

Perhaps situated cognition (Lave, 1991) can help explain first semester results. Simplistically, language is the manipulation of meaning with words. Semantic features of language become more or less salient depending on context (Caplan, 1992). Thus, language is a cognitive context. Not all contexts produce similar learning results. A hierarchy of linguistic situations exists which differ due the richness of linguistic interaction evident. Different agents serve as the primary delivery mechanism of language. From high to low level of language-rich teaching, those agents are teacher,

peers, and objects (tools). That is, in the high level of language-rich teaching, the teacher is the primary instigator of language interaction that results in students learning school concepts. In the small group and hands-on levels, peers and objects/tasks promote conceptual learning.

Anderson, Reder, and Simon (1996) argue against the requirement of complex contextual and social interaction for learning, but fail to point out any differences that might be involved in conceptual versus other types of learning. On balance, any number of studies suggest that conceptual learning is enhanced in the context of high levels of language-rich interaction (Campbell & Ramey, 1995; Hart & Risley, 1995; Landes, et al., 1995; Moje, 1995; Roth 1994, 1995).

Curriculum theory can inform these results and, in turn, these findings can mold curriculum theory. Roughly, much of curriculum theory can be placed in a continuum whose ends are teacher (top down) and student (bottom up) directed learning. Adler (1982) might represent the former and Dewey (1938) the latter. Along this spectrum, language-rich curriculum is teacher centered. That is, teachers use rules of discourse to mediate between students' utterances and intentions. Teachers shape students' conversations about concepts to be more closely aligned with teacher-experts by extensively monitoring programmed feedback sessions. For example, in the AMP journal class, the teacher monitors journal entries for accuracy as students make them. This causes students to generate linguistic representations of new concepts immediately.

Intellectual leadership is an alternative expression of teacher directed learning that informs curriculum theory. Leadership in this context means being ahead, but not out of sight, of students. Like wilderness guides, teachers lead students through engaging and demanding new intellectual territory. On the trail of learning, teachers are far enough ahead to give students direction and to help them contrast various paths along the way. Teachers are close enough to provide security and comfort in challenging sections of the journey. Clearly, classroom methods that foster students learning school concepts will

have elements of both teacher- and student-led curriculum. As Hiebert et al. (1996) suggests, it is not necessary to choose one or the other, rather to apply teaching methods that work for a particular group.

Interpreting these results relative to other studies that used the FCI places these findings in an important context. Figure 4 displays the results (in the form of two regression lines) from over 50 studies and 3000 physics students that used the FCI to monitor student achievement and teaching method effectiveness (Hake, 1994). Hake suggested two groups of studies, one that utilized "interactive engagement" and the other traditional. Teaching techniques in the interactive engagement cluster used a variety of low-tech and high-tech methods to stimulate students' linguistic representations of concepts and to compare those representations to teacher-experts. All methods required a combination of language-rich method, proper implementation, and motivated students.

Insert Figure 4 about here

Females, progressing from low to high level of language-rich teaching, were expected to improve more than males. This was due to demonstrated gender differences in spatial and mechanical abilities (Baennenger & Newcombe, 1989; Halpern, 1992), which tend to concentrate in male-taught physics classes, and verbal abilities (Hyde & Linn, 1989) which are salient to language-rich methods. Though the AMP journal females demonstrated the greatest percent gain on the FCI, unexpectedly hands-on females outscored small group females.

Dweck (1986) offers a possible explanation for this result in the *entity theory*. This theory predicts that females tend to focus on right answers because of perceived lack of ability in science. Females tend to direct their conversations towards obtaining correctness rather than conceptual understanding, unless prompted to do otherwise.

Inadvertently, the hands-on treatment required students to generate a correct theory before collecting data for the hands-on quiz. This process forced females to negotiate conceptual meaning linguistically rather than generate conversations about lab equipment and procedures exclusively. Thus, the hands-on females accessed concept formation features of language and also generated discourse about experimental tools. Females in the small group class prompted each other for right answers during group quizzes rather than challenging each other for conceptual validation.

Second semester results

All three classes were taught in the AMP journal method during the second semester. No formal instruction occurred on first semester topics during the second semester, but most first semester concepts were integral to functioning with second semester topics. For example, the concept of force is equally significant in mechanics (first semester) as in electricity and magnetism (second semester). The AMP journal was one of many optional assignments for students.

Performance on the second posttest followed the same pattern as the first posttest with the order of class means remaining AMP journal > small group > hands-on. Each class improved from posttest 1 to posttest 2 but the improvement was greater in the small group and hands-on classes. The lack of an increase, at a significance level of $p < .05$, in the class means for these two classes could have been obscured by the gender by level by test interaction. Thus, the interaction masks whether changing instructional method from low or medium level to high level language-rich teaching improves conceptual learning regardless of the time of implementation. Though suggestive, these results do not reliably argue that changing from low to high level language-rich teaching in the middle of the year improves conceptual learning. The reason why only the hands-on males and small group females improved significantly from posttest 1 to posttest 2 is

unclear. Further investigation is needed to determine the effect of switching instructional methods from lower to higher levels of language-rich teaching.

Retention is enhanced by high levels of language-rich teaching. Note that the second FCI posttest demonstrated improvement on first semester concepts even though those concepts had not been taught formally for 18 weeks. This result can be explained by the role of language in long-term semantic memory. Accessing concepts from semantic memory is different depending upon whether the concept is contained as a word (abstract concept) or by an object. Object concepts use visual identification procedures and are accessed based on physical features. Abstract concepts do not share identifiable physical features and thus are accessed by words in processes other than recognition of an object (Caplan, 1992). High level language-rich learning environments demand that students link abstract concepts to words and manipulate them at the semantic, morphological, sentence, and discourse level of language. As a result, richly nested and deeply encoded concepts become part of the intellectual automaticity of linguistic interaction. Thus, long-term memory is promoted.

Transfer is affected by the AMP journal method. An unsolicited letter to the teacher-researcher from a former student conveys this assertion.

Not only did you help me develop my skills in Chemistry and Physics, but the skills you taught me carried over into every other subject. The AMP journal helped me consolidate my note taking skills. I actually began to really think about what I was writing instead of arbitrarily and sporadically writing down facts. This process helped me, and I'm sure many others, to become a better student.

This result agrees with Anderson, Reder, and Simon (1996), who suggest that the amount of transfer depends on the teacher's ability to cue students' sense of relevance of a given concept and to engage students in multiple examples. The required question and application of the day serve this function in the AMP journal. These features of the AMP journal are developed jointly between teacher and students so as to avoid triviality and to foster a sense of common purpose between teacher and students. The greater the extent

of application and relevance as well as understandable examples that link common experience to new concepts, the greater the degree of transfer.

Limitations

Not all threats to validity were eliminated. Random assignment of subject to treatment level was not possible in this public institution. Generalizing these results should be done cautiously. FCI pretest reliability was low which clouds its effectiveness as a covariate. Some measure of remedy was afforded by obtaining similar results with the LSI as a covariate.

Not all learning is conceptual. These results apply primarily to conceptual learning and therefore should be limited to classroom situations in which conceptual content is the primary focus. Naturally, caution should be afforded when generalizing these results. It is important to replicate these results in classrooms from an array of different age, ability, and diversity backgrounds.

Implications for Teachers

The implications of this research have the greatest effect on the lives of teachers. Simplistically, a teacher's professional life can be split into an intellectual life and a daily life. Language-rich teaching can affect both.

Teachers' intellectual life

Language-rich teaching can unify the collage of learning theories prevalent today. If language-rich teaching subsumes any number of popular paradigms, then it could simplify teachers' thinking life and prevent an unbalanced application of learning theory. Teachers do not have to belong solely to the behavioral, cognitive science, or social cognition camps of learning. They can view lesson plan design through a wider and more inclusive theoretical lens. This appealing parsimony could temper the angst teachers feel

when the pendulum of reform swings by their schools. There is a certain visceral appeal to suggesting that the same language which makes us different than chimpanzees is crucial to humans learning concepts.

Three well known learning theories report ample experimental support (Hassard, 1992), yet utilize teaching methods that are language-rich to affect treatment. To illustrate, consider behavioral, cognitive science, and social cognitive views of learning. Table 6 summarizes these theories and gives examples of language-rich teaching that fall into these camps. As teachers think about helping students learn concepts, they can focus on the language-rich features of these theories and produce concomitant achievement benefits. Designing language-rich lessons draws from the primary agent for conceptual learning used in all of the other theories.

Insert Table 6 about here

As teachers create curriculum, language-rich teaching can affect their thinking. Teachers can parse the major concept words of a course or unit of study and invent lessons that cause students to use these words in as many linguistic contexts as possible. Pinker (1995) illustrates the point with three words: man, dog, and bites. Juxtaposing these words differently can either change the meaning of the sentence or make it nonsensical. Each word order conveys a conceptual context linked to humans' innate ability to learn and use language.

A physics teacher might form curriculum about the concept words force, motion, and implies. Rearranging these words in meaningful sentences produces vastly different cognizance in experts and novices. Teachers produce classroom protocol that promotes students' generation of meaning with words through discourse modeled by teachers. Through teacher demonstrations, peer discussions, and hands-on manipulation, students

use language to anchor their conceptualizations in a context generated by them in response to the intellectual leadership of a teacher-expert.

Teacher's daily life

Theories are one thing, what teachers do on Monday is another. Three suggestions focus on the daily life application of language-rich teaching. First, utilize levels of language-rich teaching in appropriate situations. For example, full-time use of the high level may lead to boredom or fatigue. Instead, use high level language-rich methods when introducing abstract concepts or those that might be particularly resilient to change. Use the medium level when students need to rehearse concepts in a social context. This helps students bridge the gap between teacher talk and student talk. Draw on the low level in situations that require procedural knowledge, recall, or categorization. Interaction with inanimate objects represents an important experiential foundation for subsequent high level language use.

Second, identify major concepts of the course and design appropriate language-rich lessons that address them. Teachers who can identify salient concepts have a much better chance of teaching them. Once teachers check their own understanding of key concept words, they can mold classroom discourse in the form of accepted conceptual understanding using language-rich techniques.

Third, many examples of successful language-rich teaching from a variety of settings have been published (Hake, 1992; Lemke, 1990; Pinkerton, 1996). Do not reinvent the wheel.

References

- Anderson, J., Reder, L., & Simon, H. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5 - 11.
- Alder, M. (1982). *The paidiea: proposal An educational manifesto*. New York: Collier.
- Baennenger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, 20, 327-343.
- Bandura, A. (1986). *Social foundations of thought and action: A social theory*. Princeton, NJ: Prentice Hall.
- Bowen, M., & Roth, W. (1995). Knowing and interacting: A study of culture, practices, and resources in a grade 8 open-inquiry science classroom guided by a cognitive apprenticeship metaphor. *Cognition and Instruction*, 13 (1), 73-128.
- Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.
- Brown, D., & Clement, J. (1991). Classroom teaching experiments in mechanics. In R. Duit, F. Goldberg, H. Niedderer (Eds.), *Proceedings of an international workshop: Research in physics learning*. Kiel, Germany: Institute fur die Padagogik der Natururssenschaften.
- Caine, R. N., & Caine, G. (1991). *Making connections: Teaching and the human brain*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Campbell, F., & Ramey, C. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, 32 (4), 743-772.
- Caplan, D. (1992). *Language: Structures, processing, and disorders*. Cambridge, MA: The MIT Press.
- Carey, S. (1985). *Conceptual change in children*. Cambridge: MIT Press.

- Chan, C., & Burtis, P. (1992). Constructive activity in learning from text. *American Educational Research Journal*, 29 (1), 97-118.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233-246.
- Chi, M., & VanLehn, K. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences*, 1, 69-105.
- Chi, M., Bassok, M., Lewis, M., Reinmann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-189.
- Chi, M., Glasser, R., & Farr, M. (1988). *The Nature of Expertise*. New Jersey: Lawrence Erlbaum Associates.
- Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3 (3), 149-211.
- Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64 (1), 1-35.
- Cohen, E. G., Lotan, R. A., & Leechor, C. (1989). Can classrooms learn? *Sociology of Education*, 62 (April), 75-94.
- Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19 (2), 237-248.
- Cook, T., & Campbell, D. (1979). *Quasi-experimental design*. Boston, MA: Houghton Mifflin Company.
- Cottle, P. D., & Lunsford, S. E. (1995). Increasing student participation in recitation. *The Physics Teacher*, 33, 23-25.
- Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology*, 8 (4), 256-262.
- Dewey, J. (1938). *Experience and education*. New York: Macmillan.

Durso, F. T., & Coggins, K. A. (1990). Graphs in the social and psychological science: Empirical contributions of Pathfinder. In R. W. Schvaneveldt (Ed.) *Pathfinder associative networks: Studies in knowledge organization* (pp. 31-52). Norwood, NJ: Ablex Publishing.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1041-1048.

Dufresne, R., Gerace, W., Leonard, W., Mestre, J., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education*, 7(2), 3-47.

Edelman, G. (1992). *Bright air, brilliant fire: On the matter of the mind*. New York: Harper Collins.

Gilbert, J. K., & Watts, D. M. (1983). Concepts, misconceptions, and alternative conceptions: Changing perspectives in science education. *Studies in Science Education*, 10, 61-98.

Glass, G. & Hopkins, K. (1984). *Statistical methods in education and psychology*. Boston: Allyn and Bacon.

Gomez, R., Hatfield, O., Houser, L. (in press). Conceptual maps and simulated teaching episodes as indicators of competence in teaching elementary mathematics. *Journal of Educational Psychology*.

Graesser, A. (1992). *Questioning mechanisms during complex learning*. Arlington, VA: Office of Naval Research.

Grumbacher, J. (1987). How writing helps physics students become better problem solvers. In T. Fulwiler (Ed.), *The Journal Book* (pp. 323-329). Portsmouth, NH: Boyton/Cook.

Gunstone, R., & Champagne, A. (1990). Promoting conceptual change in the laboratory. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum*. London: Routledge.

- Hake, R. (1992). Socratic pedagogy in the introductory physics laboratory. *The Physics Teacher*, 30 (9), 546-552.
- Hake, R. (1994). Toward mastery of Newtonian mechanics by the average student. *AAPT Announcer*, 24 (2), 55.
- Halloun, A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53, 1043-1055.
- Halloun, A., & Hestenes, D. (1987). Modeling instruction in mechanics. *American Journal of Physics*, 55, 455-462.
- Halpern, D. (1992). *Sex differences in cognitive abilities* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children..* Baltimore: Paul H. Brookes.
- Hassard, J. (1992). *Minds on science: Middle and secondary school methods*. New York: Harper Collins.
- Hestenes, D., Wells, M., & Swackhamer, S. (1992). Force concept inventory. *The Physics Teacher*, 30 (3), 141-158.
- Hiebert, J., Carpenter, T., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, 25(4), 12 - 21.
- Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education*, 22, 85-142.
- Hooper, S., & Hannafin, M. J. (1988). Cooperative CBI: The effects of heterogeneous vs. homogeneous grouping on the learning of progressively complex concepts. *Journal of Educational Computing Research*, 4, 413-424.
- Howe, A. (1996). Development on science concepts with a Vygotskian framework. *Science Education*, 80 (1), 35-51.

Howe, C., Tolmie, A., Green, K., & Makenzie, M. (1995). Peer collaboration and conceptual growth in physics: Task influences on children's understanding of heating and cooling. *Cognition and Instruction*, 13 (4), 483-503.

Hyde, J., & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104 (1), 53-69.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27(4), 664-687.

King, A. (1992). Comparison of self-questioning, summarizing and note-taking review as strategies for learning from lectures. *American Educational Research Journal*, 29 (2), 303-323.

Klopfer, L. (1990). Learning scientific inquiry in the student laboratory. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum*. London: Routledge.

Kolb, D. (1985). *Learning styles inventory*. Boston: McBer & Company.

Landes, S., Merrill, K., Merrill, D., & Reiser, B. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction*, 13 (3), 315-372.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.

Laws, P. (1991a). Workshop physics: Learning introductory physics by doing it. *Change*, 23 (4), 20-27.

Laws, P. (1991b). Calculus-based physics without lectures. *Physics Today*, 12, 24-31.

Lemke, J. (1990). *Talking science*. Norwood: Ablex Publishing.

Levene, G. (1991). *A guide to SPSS for analysis of variance*. Hillsdale, N. J.: Lawrence Erlbaum.

Lucido, H. (1992). Physics for mastery. *The Physics Teacher*, 30 (2), 96-101.

Lumpe, A. & Staver, J. (1995). Peer collaboration and concept development: Learning about photosynthesis. *Journal of Research in Science Teaching*, 32 (1), 71-98.

Lysynchuk, L. M., Pressley, M., d'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. *Reading Research Quarterly*, 24, 458-470.

Mayer, R., & Sims. V. (1994). For whom is a picture a thousand words? Extension of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86 (3), 389-401.

Mazur, E. (1993). Understanding or memorization: Are we teaching the right thing? Proceedings of the Resnick Conference on Introductory Physics Courses. (Rensselaer Polytechnical Institute, Troy, NY)

Meltzer, D., & Manivannan, K. (1996). Promoting interactivity in physics lecture classes. *The Physics Teacher*, 34 (2), 72 - 76.

Minstrell, J. (1991). Facets of student knowledge and relevant instruction. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum*. London: Routledge.

Moje, E. (1995). Talking about science: An interpretation of the effects of teacher talk in a high school science classroom. *Journal of Research in Science Teaching*, 32 (4), 349 - 371.

Newman, B. (1985). Realistic expectations for laboratory work. *Research in Science Education*, 15, 8-12.

Novak, J. D. (1990). The interplay of theory and methodology. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum*. London: Routledge.

O'Loughlin, M. (1992). Rethinking science education: Beyond Piagetian constructivism toward a sociocultural model of teaching and learning. *Journal of Research in Science Teaching*, 29 (8), 791-820.

- Pines, H. T., & West, A. L. (1985). *Cognitive structure and conceptual change*. New York: Academic Press.
- Pinker, S. (1995). *The language instinct: How the mind creates language*. New York: Harper Perennial.
- Pinkerton, K. D. (1996). Low-tech solutions, high-tech results. *The Physics Teacher*, 34 (1), 30-34.
- Pizzini, E., & Shepardson, D. (1992). A comparison of the classroom dynamic of a problem-solving and traditional laboratory model of instruction using path analysis. *Journal of Research in Science Teaching*, 29 (3), 243-258.
- Pressley, M., McDaniel, M., Tumure, J., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology*, 13 (2), 291-300.
- Prestel, B. (1993). Teaching problem solving without modeling through thinking aloud pair solving. *Journal of Science Education*, 77 (1), 83-94.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64 (4), 479-530.
- Roskes-Hofstrand, R. J., & Paap, K. R. (1986). Cognitive networks as a guide to menu organization: An application in the automated cockpits. *Ergonomics*, 29, 1301-1312.
- Roth, W. (1994). Science discourse through collaborative concept mapping: New perspectives for the teacher. *International Journal of Science Education*, 16 (4), 437-455.
- Roth, W. (1995). Affordances of computers in teacher-student interactions: The case of interactive physics. *Journal of Research in Science Education*, 32 (4), 329-347.
- Schwartz, D. L., Black, J. D., & Strange, J. (1991, April). *Dyads have a fourfold advantage over individuals in inducing abstract rules*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Scott, H., Asoko, H., & Driver, R. (1991). Teaching for conceptual change: A review of strategies. In R. Duit, F. Goldgerg, H. Niedderer (Eds.), *Physics learning: Theoretical issues and empirical studies*. Institute for Science Education, University of Bremen, Germany.

Sere, M. (1991). Learning by giving and receiving explanations. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum*. London: Routledge.

Shachar, H. & Sharan, S. (1994). Talking, relating, and achieving: Effects of cooperative learning and whole-class instruction. *Cognition and Instruction*, 12 (4).

Simpson, M. L., Olejnik, S., Tam, A., & Supattathum, S. (1994). Elaborative verbal rehearsals and college students' cognitive performance. *Journal of Educational Psychology*, 86 (2), 267-278.

Slamecka, N., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning and Memory*, 4 (6), 592-604.

Schvaneveldt, R. (ed.) (1990). *Pathfinder networks: Studies in knowledge organization*.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8 (4), 351-362.

Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics*. New York: Harper Collins.

Thornton, R., & Sokoloff, D. (1990). Learning motion concepts using real-time microcomputer-based laboratory tools. *American Journal of Physics*, 58, 858-867.

Tobin, K. (1990). Research on science laboratory activities: In pursuit of better questions and answers to improve learning. *School Science & Mathematics*, 90, 403-418.

Trudge, J. (1990). Vygotsky: The zone of proximal development and peer collaboration: Implications for classroom practice. In L. Moll (Ed.), *Vygotsky and*

education: Instructional implications and applications of sociohistorical psychology.
New York: Columbia University Press.

Vygotsky, L. S. (1978). *Mind is society*. Cambridge, MA: Harvard University Press.

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21-39.

Webb, N. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22 (5), 366-389.

Whelan, M. T. (1994). Scientific and everyday concept formation: An application of L.S. Vygotsky's work to the speech theory of human communication. Unpublished doctoral dissertation, University of Denver.

Wilson, J. (1994). The CUPLE physics studio, *The Physics Teacher*, 32, 518-523.

Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist*, 24 (4), 345-376.

Table 1

Summary of Demographic Data

Category	Level					
	AMP Journal		Small Group		Hands-on	
Gender	m(22)	f(9)	m(12)	f(11)	m(15)	f(15)
Years math	2.8		2.8		3.0	
Years science	2.7		2.6		2.7	
GPA	3.3		3.2		3.2	
Age(years)	17.7		17.6		17.7	

Note. Numbers in parentheses are exact counts. All other values are means.

Table 2

Descriptive Statistics for ANCOVA of ECI Posttest with Pretest Covariate

Cell	n	Raw <u>M</u>	Adjusted <u>M</u>	<u>SD</u>	.95 CI
AMP males	19	21.8	20.1	3.42	± .93
AMP females	9	17.9	18.6	5.09	±1.18
Small group males	13	20.8	19.4	4.10	±1.05
Small group females	10	13.5	15.6	1.78	±1.29
Hands-on males	14	17.6	16.4	3.37	± .98
Hands-on females	14	16.1	17.5	2.43	±1.01

Table 3

Analysis of Covariance for Language-rich Teaching as Measured by FCI

Source	df	η^2	power	E
				FCI-post with pretest covariate
Gender	1	.025	.27	1.88
Period	2	.080	.58	3.12*
Gender x Period	2	.073	.54	2.84‡
Within subjects error	72	na	na	(13.3)
Contrast 1: AMP vs (SG+HO)	1	.086	.73	6.81*
Contrast 2: Gender by Level (SG vs HO)	1	.064	.59	4.88*

Note. Value in parentheses is the mean square error.

* $p < .05$ ‡ $p = .06$

Table 4

Descriptive Statistics for ANCOVA of NETSIM Post Measure with Pre Measure
Covariate

Cell	n	Raw <u>M</u>	Adjusted <u>M</u>	<u>SD</u>	.95 CI
AMP males	19	.159	.144	.105	±.022
AMP females	9	.130	.131	.105	±.031
Small group males	13	.126	.117	.082	±.025
Small group females	10	.047	.052	.097	±.035
Hands-on males	14	.065	.068	.081	±.025
Hands-on females	14	.072	.085	.062	±.025

Table 5

Analysis of Covariance for Language-rich Teaching as Measured by NETSIM

Source	df	η^2	power	E
				Posttest with pretest as covariate
Gender	1	.012	.17	.89
Period	2	.083	.60	3.24*
Gender x Level	2	.034	.27	1.27
Within subjects error	72	na	na	(.01)
Contrast 1: AMP vs. (SG+HO)	1	.087	.73	6.85*
Contrast 2: Gender x Level	1	.028	.29	2.04
(SG vs HO)				

Note. Value enclosed in parentheses is the mean square error.

* $p < .05$

Table 6

Summary of Learning Theories with Language-rich Teaching Examples

Theory	Metaphor	Operative Word	Model	Example from Language-rich
Behavioral	mind is like a muscle	reinforcement	directed learning	teacher reinforcement of correct concept word use in AMP journal
Cognitive science	mind is like a computer	process	constructivism, inquiry learning	calling up students' existing concepts in verbal format to process conceptual change
Social cognition	mind is like a coffee house	interaction	cooperative learning	turn and talk interaction episodes, application of the day

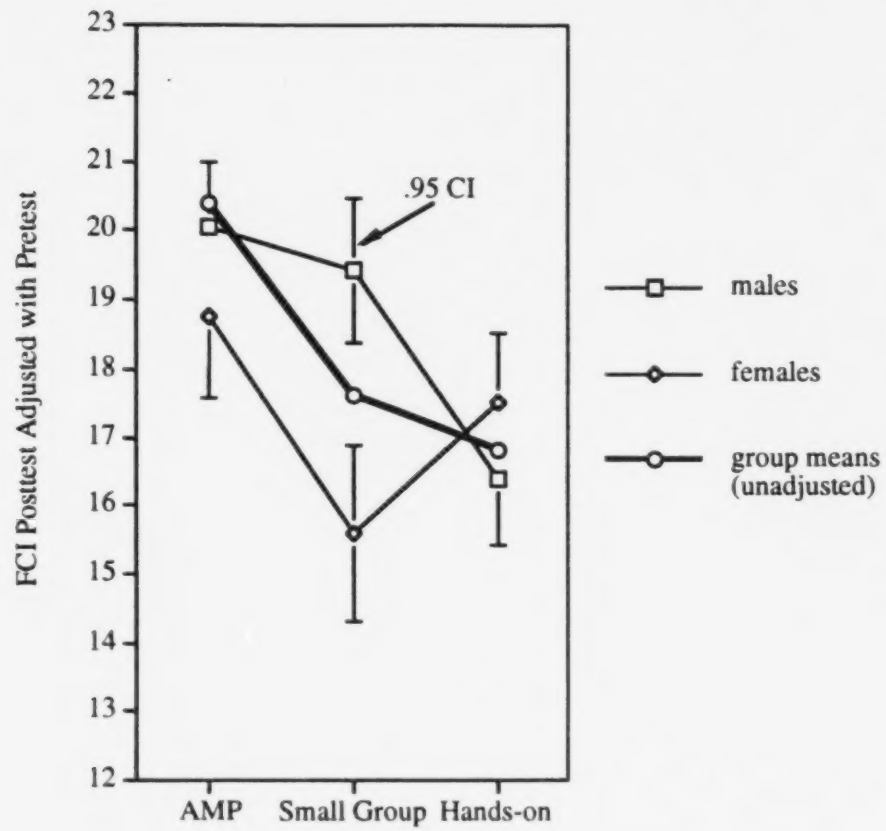


Figure 1.

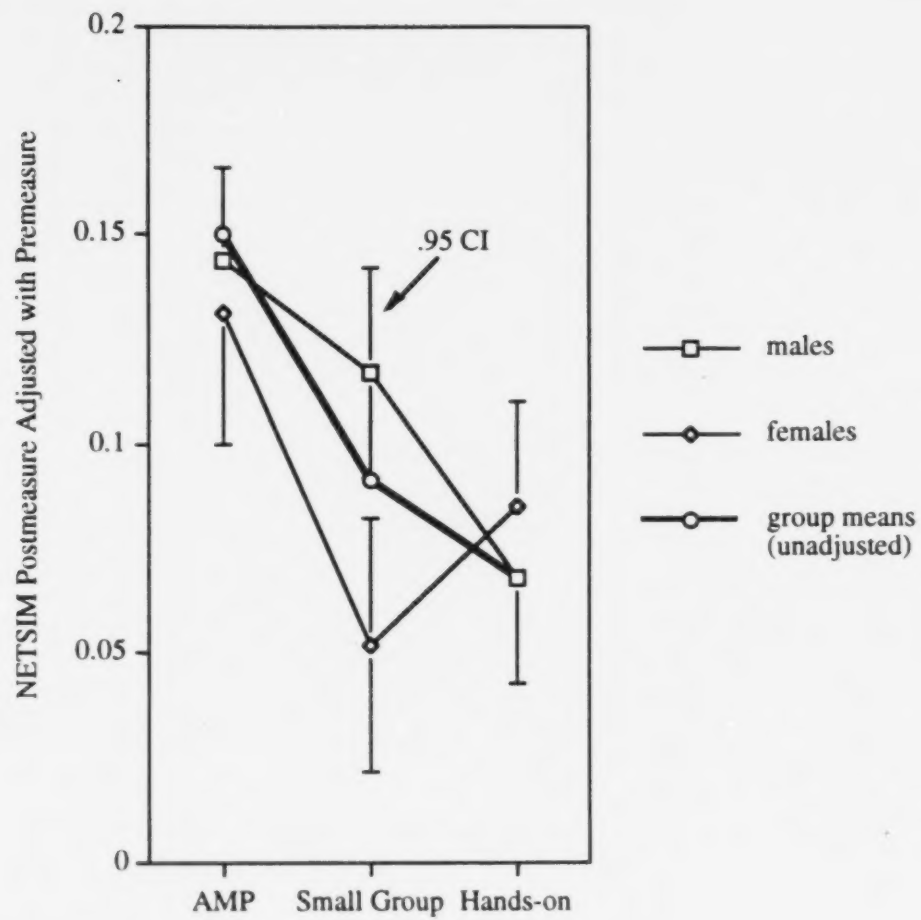


Figure 2.

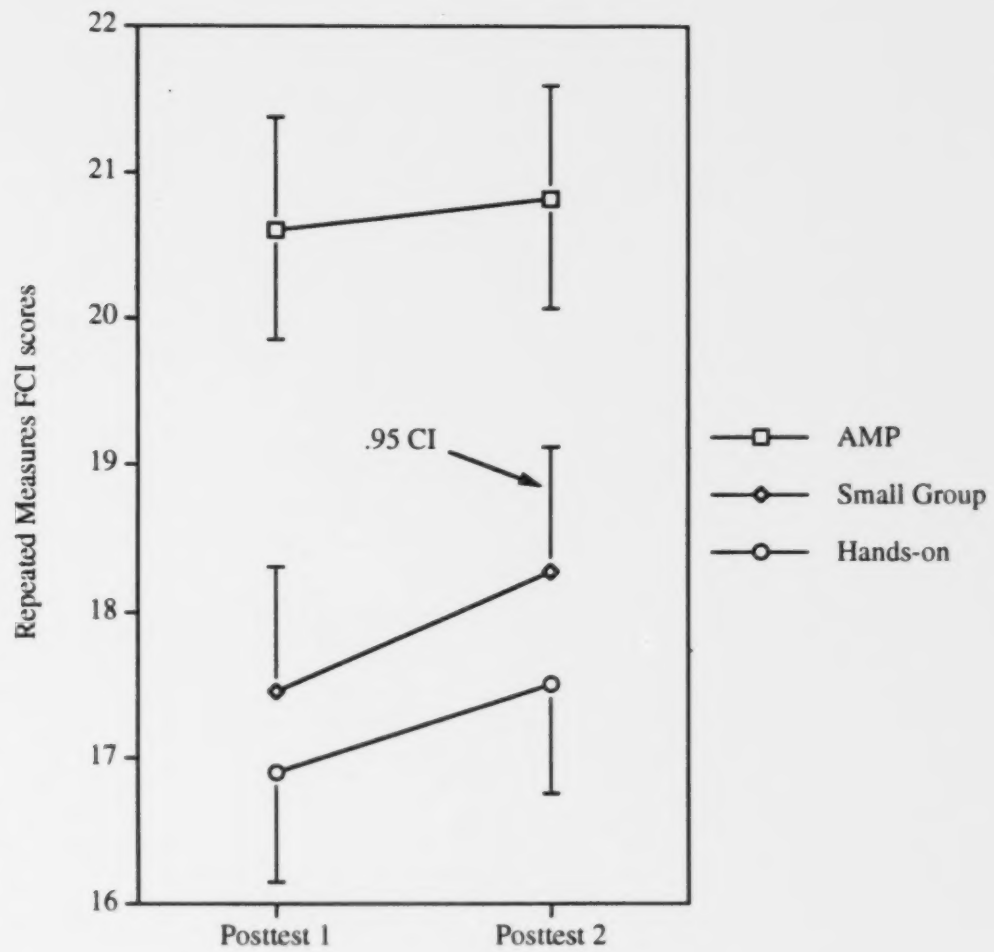


Figure 3.

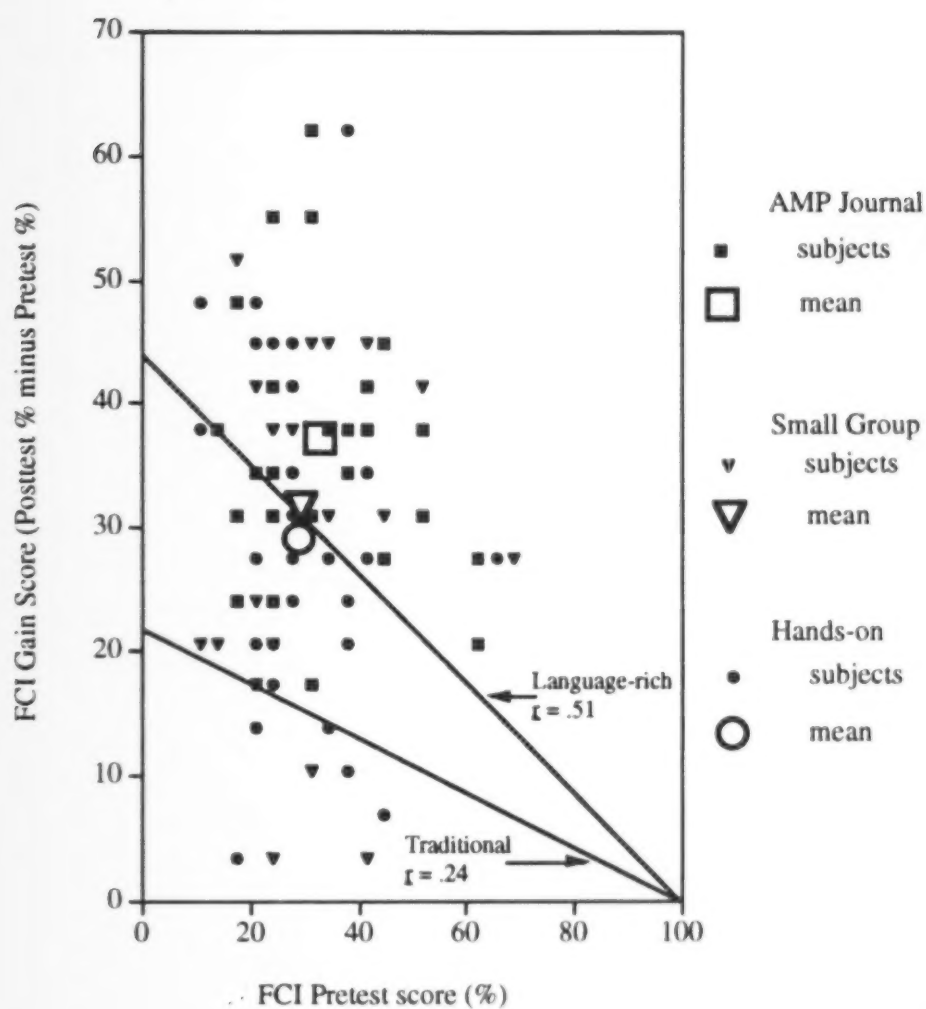


Figure 4.

Captions

Figure 1. FCI posttest scores adjusted with pretest as covariate for three levels of language-rich teaching.

Figure 2. NETSIM postmeasures adjusted with premeasures as covariate for three levels of language-rich teaching.

Figure 3. Raw class means of repeated measures FCI posttests. Tests given 18 weeks apart.

Figure 4. Comparison of Hake regression lines with gain scores of AMP journal, small group, and hands-on classes.

Appendix
Sample Lessons

Active Mental Processing Journal

sample lesson

Previous Knowledge: Students know the definition of acceleration, velocity, displacement and vectors. They have introductory skills with vectors.

Concept to be Taught: Independent nature of vertical and horizontal motion.

Script: Day 1

Each year the US. supplies hundreds of tons of food to starving people, sometimes in war torn areas. Of necessity, airplanes drop this relief. The pilots must do this accurately lest they invite disaster. Luckily, physics concepts are part of the answer.

Let me demonstrate a concept that will help you drop the food exactly on target. (I hold up a demonstration device that supports two identical steel balls. It has a spring loaded rod that will hit one ball horizontally at the same instant that it releases the other ball to fall straight down.) Please sketch this device in your AMP journal. (Give students sketching time. Walk around the room and monitor the drawings.) The sketches look good. Now elaborate your drawings by labeling the following important features; rod, spring, ball that falls straight down, ball that is hit horizontally, and both balls begin moving at the same time.

Please predict what you will hear as the balls hit the ground. Will you hear this (Clap hands once.) or this (Clap hands twice.)? Write your prediction in your journal and don't forget to say why (I walk about the room to monitor students answers.).

Now turn and talk to your neighbor. Read your prediction to them and explain your ideas. When you are done, allow them to do the same. (I listen to the groups.)

I am going to perform the demonstration. Please write exactly what you hear and see in your AMP journal. (The balls hit at the same time. Allow time to write observations.)

Why did this happen? (Someone usually says that it is due to the balls having the same vertical acceleration. What are the characteristics of the horizontal motion? (A

student usually says that it constant because of inertia.) When does the horizontal motion of the ball hit sideways stop? (Students respond, when the vertical motion stops.) Now while it is fresh in your head, write a note to yourself in your journal. There are at least three important points to make. Please do it in this form:

What I saw	What it means
1.	1.
2.	2.
3.	3.

(I walk around the room to monitor the writing.)

Let's relate this to a real world situation. Remember the relief food that I mentioned at the first of the class period? In your AMP journals, draw a sketch of a plane carrying food under its fuselage and traveling at a constant velocity and height above the ground. At what point in the flight path would the pilot release the food in order for it to hit a target? Include the path of the food and the position of the plane at the moment the food hits the target. (Walk around and monitor the answers.)

We have only three minutes remaining in the period. Please write your question of the day in your journals.

Day 2

Please take out your AMP journals and review your sketch of the food's path. Write yourself a note explaining whether the food will hit the ground behind, directly below, or ahead the plane. (Monitor their responses) Turn and talk to your neighbor. Explain to him or her what you have written. Please allow for equal time.

Let's relate the two ball demonstration to the food drop. We'll do it this way. Sketch a graph of distance versus time, velocity versus time, and acceleration versus time for the straight down ball and the vertical motion of the food. (Walk around the room and help to correct mistakes.)

Please go beyond answering my question and elaborate your notes by writing a caption of explanation on each graph. Tell yourself the meaning of each graph. (Monitor this activity.)

Now, do the same thing for the horizontal motion of the food and the ball hit horizontally. Don't forget to elaborate your sketches.

Let's summarize what we've learned. Horizontal and vertical motion are independent of one another. That is, the motion of a projectile hit horizontally close to the surface of the earth can be analyzed in two components, horizontal and vertical. What governs the horizontal motion is inertia and it is completely independent of the constant force of gravity which governs the vertical motion.

Oh look! We have only five minutes left in the period. Please write your application of the day while it is fresh in your mind.

Small Group
sample lesson

Previous Knowledge: Students know the definition of acceleration, velocity, displacement and vectors. They have introductory skills with vectors.

Concept to be Taught: Independent nature of vertical and horizontal motion.

Script: Day 1

Each year the US. supplies hundreds of tons of food to starving people, sometimes in war torn areas. Of necessity, airplanes drop this relief. The pilots must do this accurately lest they invite disaster. Luckily, physics concepts are part of the answer.

Let me demonstrate a concept that will help you drop the food exactly on target. (I hold up a demonstration device that supports two identical steel balls. It has a spring loaded rod that will hit one ball horizontally at the same instant that it releases the other ball to fall straight down.) Notice that the device will release one ball to fall straight down at exactly the same instant it hits the other ball horizontally.

Please predict what you will hear as the balls hit the ground. Will you hear this (Clap hands once.) or this (Clap hands twice.)? Let's let nature tell us the answer. (I perform the demonstration. Both balls hit at the same time.)

Why did this happen? (Someone usually says that it is due to the balls having the same vertical acceleration. What are the characteristics of the horizontal motion? (A student usually says that it constant because of inertia.) When does the horizontal motion of the ball hit sideways stop? (Students respond, when the vertical motion stops.)

Let's relate this to a real world situation. Remember the relief food that I mentioned at the first of the class period? At what point in the flight path would the pilot release the food in order for it to hit a target? You are right! The food hits exactly under the plane because the food and the food have the same horizontal motion until the food hits the ground. The amount of time the food is in the air depends on the release altitude

and the acceleration of gravity. Therefore the food can travel forward only for the amount of time that the food is falling toward the earth.

Let's relate the two ball demonstration to the food drop. We'll do it this way. I am going to use our graphing skills to help explain why the balls hit at the same instant and why the food hits directly under the plane. (I draw and explain the shapes of displacement versus time, velocity versus time, and acceleration versus time in both horizontal and vertical components of motion for the two balls and the food.)

Let's summarize what we've learned. Horizontal and vertical motion are independent of one another. That is, the motion of a projectile hit horizontally close to the surface of the earth can be analyzed in two components, horizontal and vertical. What governs the horizontal motion is inertia and it is completely independent of the constant force of gravity which governs the vertical motion.

Okay, tomorrow we will have a quiz. You will have one class period to answer questions on what you have learned today.

Day 2

Today is your quiz. Remember that you may work in your small groups on the answers only after everyone has attempted all the questions. When you do start helping each other, please remember to use the talk aloud methods that I taught you. Good luck and here is your quiz. (I pass out the following questions on a separate piece of paper.)

1. How many "clicks" do you hear when the coins hit the floor? Please explain why this happens.
2. Sketch the horizontal and vertical velocity versus time graphs for each coin. Explain any differences and similarities.
3. Suppose an airplane is flying at a constant horizontal velocity with a large lead ball attached to its bottom. The ball suddenly falls. Will the lead ball fall behind, directly beneath, or in front of the plane? Provide a sketch and include the path of the ball as well as the position of the plane exactly at the instant the ball hits the ground.